What is claimed is:

1. A method for the synthesis of photo-realistic animation of an object using a unit selection process, comprising the steps of:

   a) creating a first database of image samples showing the object in a plurality of appearances;

   b) creating a second database of visual features for each image sample of the object;

   c) creating a third database of non-visual characteristics of the object in each image sample;

   d) obtaining for each frame in the plurality of N frames of the animation, a target feature vector comprised of visual features and/or non-visual characteristics.

   e) performing, for each frame in the plurality of N frames of the animation, a unit selection process to select candidate image samples from the first database using a combination of visual features from the second database and associated object characteristics from the third databases to compare with the target feature vector.

   f) compiling the selected candidates to form the photo-realistic animation.

2. The method as defined in claim 1 wherein the visual features of the second database are extracted from intermediate images representing normalized sub-parts of the object obtained from the image sample of the first database.

3. The method as defined in claim 2 wherein the normalization comprise the steps of:

   a) calculating the pose of the object (rotation angles and position in 3D space) as it appears on an image sample of the first database;

   b) reprojecting the object onto an intermediate image using a normalized pose.

4. The method as defined in claim 3 wherein the pose of the object is calculated using a set of at least four 3D object points and their corresponding image projection and applying standard pose estimation algorithms.

5. The method as defined in claim 3 wherein reprojection step is performed as follows:

   a) 3D quadrilaterals defining the overall shape of the object are projected on the image using the object's calculated pose, marking 2D quadrilateral boundaries;

   b) the same quadrilaterals are projected onto an intermediate image using the standard pose, marking a second set of 2D quadrilaterals.

   c) a standard quadrilateral-to-quadrilateral mapping is performed for each quadrilateral in the object from the image sample to the intermediate, normalized image.

6. The method as defined in claim 2 wherein the features include the projections of the normalized sub-part image onto a subset of its principal components. The principal components being calculated from the entire set of available normalized sub-part images, using standard PCA (principal component analysis).

7. The method as defined in claim 2 wherein the visual features include a wavelet decomposition of the images. Each image is transformed with a wavelet transform. A subset of the wavelet coefficients is selected as feature vectors for the images.

8. The method as defined in claim 2 wherein the visual features include a projection onto a set of selected template images. Each image is overlaid onto a set of template images and a pixel-by-pixel multiplication is calculated (alternatively a pixel-by-pixel difference). The coefficients calculated in this way represent the feature vectors for the images.

9. The method as defined in claim 6 wherein PCA is performed on subsampled and cropped images of the normalized image samples.

10. The method as defined in claim 6 wherein PCA is performed on luminance images of the normalized image samples.

11. The method as defined in claim 1 wherein in performing the unit selection process, the following steps are performed:

a) for each frame, a number of candidates image samples from the first database are selected based on the target feature vector and object characteristics from the third database;

b) for each pair of candidates of two consecutive frames, a concatenation cost is calculated from a combination of visual features from the second database and object characteristics from the third database;

c) a standard Viterbi search is performed to find the least expensive path through the candidates accumulating the target and concatenation costs.

12. The method as defined in claim 11, step b) wherein, the concatenation cost is given by the Euclidian distance in the space of visual features between two candidates.

13. The method as defined in claim 12 wherein an additional concatenation cost g is calculated from the respective recording timestamps of the image samples $u1$, $u2$ using the following formula:

$$g(u1,u2)=\begin{cases} 0 \text{ when } fr(u1)-fr(u2)=1 \wedge seq(u1)=seq(u2) \\ w_1 \text{ when } fr(u1)-fr(u2)=0 \wedge seq(u1)=seq(u2) \\ w_2 \text{ when } fr(u1)-fr(u2)=2 \wedge seq(u1)=seq(u2) \\ \ldots \\ w_{p-1} \text{ when } fr(u1)-fr(u2)=p-1 \wedge seq(u1)=seq(u2) \\ w_p \text{ when } fr(u1)-fr(u2)\geq p \vee fr(u1)-fr(u2)<0 \\ \qquad\qquad \vee seq(u1)\neq seq(u2) \end{cases} \quad \text{where}$$

$0<w_1<w_2<\ldots<w_p$, $seq(u) = recorded\_sequence\_number$ and $fr(u) = recorded\_frame\_number$.

14. The method as defined in claim 1 wherein the animation is a talking-head animation, the object a human head, the first database stores sample images of a face that speaks, the second database stores associated facial visual features and the third database stores acoustic information for each frame in the form of phonemes.

15. The method as defined in claim 4 wherein the pose of the object is calculated using the position of the inner and outer corners of the left and right eye and the two nostrils.

16. The method as defined in claim 14 wherein visual features are extracted from normalized images of the mouth area including lips, chin and cheeks.

17. The method as defined in claim 16 wherein the visual features of the normalized mouth samples include projections onto a set of principal components. The principal components being calculated using PCA on the entire database of normalized mouth samples.

18. The method as defined in claim 16 wherein the visual features of the normalized mouth samples include shape and position of the outer and inner lip contour, of the upper and lower teeth and of the tongue.

19. The method as defined in claim 11, step a) wherein, the target cost is calculated by the following steps:

a) defining a phonetic context by including in the cost calculation ***nl*** frames left of the current frame and ***nr*** frame right of it.

b) obtaining a target phonetic vector for each frame t, the target feature vector described as $\boldsymbol{T(t)} = \{ph_{t-nl}, ph_{t-nl-1}, \ldots, ph_{t-1}, ph_t, ph_{t+1}, \ldots, ph_{t+nr-1}, ph_{t+nr}\}$, where ph_i is the phoneme being articulated at frame i.

c) defining a weight vector $\boldsymbol{W(t)} = \{w_{t-nl}, w_{t-nl-1}, \ldots, w_{t-1}, w_t, w_{t+1}, \ldots, w_{t+nr-1}, w_{t+nr}\}$;

d) defining a phoneme distance matrix **M[p1,p2]** that gives the distance between two phonemes;

e) getting a candidate's phonetic vector from the third database $\boldsymbol{U(u)} = \{ph_{t-nl}, ph_{t-nl-1}, \ldots, ph_{t-1}, ph_t, ph_{t+1}, \ldots, ph_{t+nr-1}, ph_{t+nr}\}$;

f) computing the target cost TC, using the following:

$$TC(t,u) = \frac{1}{\sum_{i=-nl}^{nr} w_{t+i}} \sum_{i=-nl}^{nr} w_{t+i} \cdot M(T_{t+i}, U_{u+i}),$$

20. The method as defined in claim 19, step c) wherein the elements of the weight vector are calculated using the following equation: $w_i = e^{-\alpha|t-i|}$

21. The method as defined in claim 19, step d) wherein the phoneme distance matrix **M** is populated using similarity between their visemic representation.